

109

Data Documentation

Anonymisierungskonzept für den Scientific Use File zur Studie „Erwartungen an Wirtschaft und Gesellschaft“, Zweite Welle (LINOS-2)

IMPRESSUM

DIW Berlin, 2024

DIW Berlin

Deutsches Institut für Wirtschaftsforschung

Mohrenstr. 58

10117 Berlin

Tel. +49 (30) 897 89-0

Fax +49 (30) 897 89-200

www.diw.de

ISSN 1861-1532

All rights reserved.

Reproduction and distribution

in any form, also in parts,

requires the express written

permission of DIW Berlin.

Data Documentation 109

Jule Adriaans¹

Tobias Gebel²

Anonymisierungskonzept für den Scientific Use File zur Studie „Erwartungen an Wirtschaft und Gesellschaft“, Zweite Welle (LINOS-2)

Berlin, April 2024

¹ Universität Bielefeld, Fakultät für Soziologie, jule.adriaans@uni-bielefeld.de

² DIW Berlin, tgebel@diw.de

Inhalt

1. Ausgangslage.....	1
2. Kurzbeschreibung der Studie	1
3. Erstellung des Scientific Use Files	2
3.1. Angewandte Regeln zum Zweck der faktischen Anonymisierung	2
3.2. Löschung von Merkmalen	3
3.3. Aggregation von Merkmalen.....	3
3.4. Änderungen im Einzelnen	4
4. Datenzugang	6

1. Ausgangslage

Bei der Studie „Erwartungen an Wirtschaft und Gesellschaft“ handelt es sich um eine bundesweite Erwerbstätigenbefragung zum Thema der Herausbildung von Gerechtigkeitseinstellungen über den Lebensverlauf. Die zweite Welle (LINOS-2) der als Wiederholungsbefragung angelegten Studie, wurde 2017 im Rahmen des DFG-geförderten Projektes „Strukturelle Bedingungen von Gerechtigkeitseinstellungen über den Lebensverlauf“ durchgeführt. Grundlage für die zweite Welle sind alle Teilnehmenden aus der ersten Welle (LINOS-1), die einer erneuten Kontaktierung zugestimmt haben. Grundlage der LINOS-1 Stichprobe sind Meldedaten der Bundesagentur für Arbeit.³ Eine detaillierte Beschreibung des Studiendesigns und der Stichprobenziehung für LINOS-1 findet sich in Sauer & Valet (2014).⁴

Gemäß den Prinzipien guter wissenschaftlicher Praxis⁵ werden die Befragungsdaten durch das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am DIW Berlin in anonymisierter Form als Scientific Use File (SUF) weiteren WissenschaftlerInnen zur Verfügung gestellt.

Das vorliegende Dokument beschreibt das Konzept zur Erstellung des Scientific Use Files.

2. Kurzbeschreibung der Studie

Ziel der Studie „Erwartungen an Wirtschaft und Gesellschaft“ ist es zu untersuchen, warum Menschen bestimmte Gerechtigkeitseinstellungen haben. Als entscheidend werden dabei die u.U. wechselnden sozialen Kontexte – insbesondere Haushalte, soziale Netzwerke oder Betriebe – angesehen, in denen Personen über ihren Lebensverlauf eingebunden sind. Denn soziale Kontexte vermitteln Möglichkeiten des sozialen Vergleichs und des sozialen Lernens, die für die Ausbildung von Gerechtigkeitseinstellungen entscheidend sind.

Die Bruttostichprobe der zweiten Befragungswelle (LINOS-2) sind die insgesamt 3.607 LINOS-1 Befragten, die in der ersten Befragungswelle ihre Einwilligung zur Wiederkontaktierung erteilt haben. Auf Basis dieser Ausgangsstichprobe konnten insgesamt 2.741 Interviews in LINOS-2 realisiert werden.

Die Erhebung der Daten erfolgte auf drei Arten:

1. Ohne Interviewer: Papierfragebögen (PAPI, paper and pencil interview)
2. Ohne Interviewer: Web-Interviews (CAWI, computer assisted web interview)
3. Mit Interviewer: Persönliche Interviews (CAPI, computer assisted personal interview)

LINOS-2 Befragte wurden im gleichen Erhebungsmodus kontaktiert wie in LINOS-1. Eine detaillierte Dokumentation des Studiendesigns, der Feldarbeit und der erhobenen Merkmale findet sich in Adriaans et al. (2019).⁶

³ Die Ziehung der LINOS-1 Stichproben erfolgte aus denjenigen Personen, die am 31.12.2011 sozialversicherungspflichtig beschäftigt sowie im Alter zwischen 19 und 59 Jahren waren.

⁴ Sauer, Carsten & Valet, Peter (2014): „*Erwartungen an Wirtschaft und Gesellschaft. Feldbericht und Codebuch zur Erwerbstätigenbefragung. Erste Welle*“ SFB 882 Technical Report Series 10. DFG Research Center (SFB) 882 From Heterogeneities to Inequalities, Bielefeld.

⁵ Betancort, Noemi; Bongartz, Elke C.; Dörrenbächer, Nora; Goebel, Jan & Kaluza, Harald; Siegers, Pascal (2020): „White Paper on implementing the FAIR principles for data in the Social, Behavioural, and Economic Sciences“ *RatSWD Working Paper Series 274*. URL: https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_274.pdf

⁶ Adriaans, Jule; Eisnecker, Phillip; Hülle, Sebastian; Klassen, J.; Liebig, Stefan; Valet, Peter (2019): „Erwartungen an Wirtschaft und Gesellschaft – Legitimation of Inequality over the Life-Span. Feldbericht und Codebuch zur zweiten Welle (LINOS-2).“ *DIW Data Documentation 97*. Deutsches Institut für Wirtschaftsforschung DIW Berlin.

3. Erstellung des Scientific Use Files

Für die Erstellung des Scientific Use Files wird das Konzept der faktischen Anonymisierung angewandt. Dabei bedeutet faktisch anonym, dass eine Deanonymisierung einzelner im Datensatz enthaltener Individuen nur mit unverhältnismäßig hohem Aufwand an Zeit, Kosten und Arbeit möglich ist und somit die Kosten der Deanonymisierung den Nutzen derselben übersteigt. Anwendung finden dafür die Regeln von Müller et al. (1991)⁷ zur faktischen Anonymisierung von Mikrodaten.

Da der Informationsgehalt für die Genauigkeit der Analysen wichtig ist, sollen die Daten so weit wie möglich erhalten bleiben. Bei der Erstellung des Scientific Use Files kommen deshalb ausschließlich die Regeln der Löschung einzelner Merkmale und Aggregationen von sensiblen Informationen zum Einsatz.

3.1. Angewandte Regeln zum Zweck der faktischen Anonymisierung

Das Ausgangsmaterial der Befragung „Erwartungen an Wirtschaft und Gesellschaft“ weist bereits ein geringes Reidentifikationsrisiko der Befragten auf: Zum einen sind die individuellen Auswahlwahrscheinlichkeiten, die sich aus der breiten Auswahlgrundlage vieler Millionen Beschäftigter und der realisierten Stichprobe von nur 2.741 Befragten ergeben, sehr gering. Zudem wird auf eine regionale Tiefengliederung weitgehend verzichtet, es werden lediglich zwei Regionen unterschieden: Ost und West.

Die Anonymisierung erfolgt über die Löschung von Merkmalen und über die Vergrößerung vorhandener Antwortkategorien. Subjektive Angaben und Einschätzungen stellen die Mehrzahl der Merkmale dar. Diese sind im Normalfall als unproblematisch hinsichtlich des Risikos der Reidentifikation zu betrachten und können bleiben erhalten. Darunter fallen auch alle im Rahmen der faktoriellen Surveys⁸ und des Choice Sets⁹ gemachten Angaben. Hingegen können durch die Kombination von soziodemografischen Merkmalen oder harten Fakten Individuen direkt identifiziert werden, so dass hier eine Anonymisierung durch Löschung oder Aggregation notwendig ist.

Die Löschung oder Aggregation von Merkmalen ist erforderlich, wenn zu erwarten ist, dass in der Grundgesamtheit eine geringe Zellbesetzung vorliegt. Schlüsse über die Zellbesetzung der Grundgesamtheit werden sowohl auf Basis von Auszählungen dieser Befragung als auch auf Basis anderer Datenquellen gemacht.

⁷ Müller, Walter; Blien, Uwe; Knoche, Peter & Wirth, Heike (1991): *Die faktische Anonymität von Mikrodaten*. Stuttgart: Metzle-Poeschel.

⁸ Faktorielle Surveys bestehen aus fiktiven Beschreibungen von Personen oder Objekten (sog. Vignetten). Diese werden durch verschiedene Attribute (Dimensionen) beschrieben, die in ihren Ausprägungen (Levels) variieren. In dieser Studie wurden zwei verschiedene Vignettenmodule eingesetzt: Einkommensvignetten und Arbeitsangebotsvignetten. Es folgen zwei Beispiele, bei denen die variierten Attribute fett markiert sind.

Einkommensvignette: Eine **30-jährige Frau** ist als **Hilfsarbeiterin** beschäftigt. Sie arbeitet als **Leiharbeiterin** in einem Betrieb und erbringt dort **überdurchschnittliche** Leistung. Die Arbeitslosigkeit in ihrem Beruf ist **hoch**. Ihr monatliches Bruttoeinkommen beträgt **1500 Euro**.

Arbeitsmarkt vignette: Eine **30-jährige Frau** mit **abgeschlossener Berufsausbildung** ist nach eigener **Kündigung** arbeitslos. Sie hat sich kürzlich in einem Betrieb als **Produktionsleiterin** beworben. Von diesem Betrieb erhält sie jetzt ein Angebot für einen **unbefristeten** Arbeitsvertrag mit einem Einkommen, das deutlich niedriger ist als sonst üblich. Es gibt eine **Gleitzeitregelung** mit **Arbeitszeitkonto**. Die Belastungen am Arbeitsplatz sind **durchschnittlich hoch**.

Für jedes Fallbeispiel wurde erfragt, ob es als gerecht oder ungerecht beurteilt wird.

⁹ Das Discrete Choice Set beschreibt in tabellarischer Form drei Arbeitsverträge mit unterschiedlichen Rahmenbedingungen (Dimensionen). In jedem Arbeitsvertrag wurden die Dimensionen Fort- und Weiterbildungsmöglichkeiten, Lohnhöhe, Vertragslaufzeit, Vereinbarkeitsangebote für Pflege und Familie und das Ansehen des Unternehmens in ihren Ausprägungen variiert. Die Zielpersonen sollten sich jeweils für einen der drei Arbeitsverträge entscheiden. Insgesamt wurden zwölf verschiedene Sets mit je drei Arbeitsverträgen konzipiert und jedem Befragten eines der zwölf Sets zufällig zur Beantwortung zugewiesen.

3.2. Löschung von Merkmalen

Die systemfreie Interviewer-ID (bi1) bleibt erhalten. Die Merkmale zur Interviewsituation bei persönlichen Interviews (bi2 – bi11) bleiben ebenfalls erhalten. Sie stellen kein Reidentifikationsrisiko dar, enthalten aber Informationen zur Datenqualität und erlauben dadurch eine für die Auswertung mitunter wichtige Bewertung unplausibler Antworten.

Alle Merkmale, die Klartextangaben enthalten, werden gelöscht.

3.3. Aggregation von Merkmalen

Soziodemographische Merkmale stellen ein Reidentifikationsrisiko dar, wenn sie Merkmalsausprägungen enthalten, die leicht beobachtbar sind und in der Grundgesamtheit selten vorkommen. Die folgenden Merkmale werden aggregiert:

- Alter: Es liegen das Geburtsjahr und der Geburtsmonat des Befragten vor. Der Geburtsmonat wird gelöscht, Geburtsjahre vor 1954 und nach 1993 werden zusammengefasst.
- Haushaltszusammensetzung, Kinder- und Geschwisterzahl: Bei der Zahl der Personen im Haushalt, der Zahl der Kinder und der leiblichen Kinder sowie der Zahl der Geschwister des Befragten und seiner Position in der Reihenfolge der Geschwister werden hohe Werte zusammengefasst, um die Reidentifikation von Großfamilien zu verhindern.
- Zur aktuellen Erwerbssituation liegen detaillierte Angaben für den Befragten und dessen Partner vor. Die Items entsprechen weitgehend dem Mikrozensus, so dass anhand der Fallzahlen des Mikrozensus auf die Fallzahl in der Grundgesamtheit geschlossen werden kann. Einzelne Kategorien zur Erwerbssituation werden jeweils so zusammengefasst, dass sie nicht feiner sind als im Scientific Use Files des Mikrozensus 2011.
- Sofern der Befragte Vorgesetzter ist, wird der nach der Anzahl der beaufsichtigten Personen befragt. Zum Schutz von Top-Managern werden hohe Werte zusammengefasst.
- Angaben zur aktuellen Tätigkeit des Befragten wurden anhand von offenen Fragen detailliert erfasst und anschließend in Berufsklassifikationen eingeordnet. Im SUF bleibt lediglich zwei Merkmal erhalten: Beruf nach ISCO-Klassifikation als 4-Steller. Dies entspricht dem u.a. dem Vorgehen des Sozio-ökonomischem Panel.
- Erwerbshistorie: Für nicht-erwerbstätige Befragte und deren Partner wurden jeweils der Monat und das Jahr der letzten Erwerbstätigkeit erfragt. Diese Information wird durch das Löschen des Monats vergrößert.
- Da die Daten zur Untersuchung der Herausbildung von Gerechtigkeitseinstellungen verwendet werden sollen, sind die Einkommensangaben des Befragten besonders wichtig und sollen möglichst wenig vergrößert werden. Es werden das persönliche Brutto- und Nettoeinkommen sowie das Nettohaushaltseinkommen erfragt. Die Angaben werden auf einen Maximalwert von 10.000 € zensiert. Neben den realen Einkommen werden auch die jeweils als gerecht empfundenen Einkommen erfragt. Dabei handelt es sich um subjektive Einschätzungen, so dass diese Angaben unverändert erhalten bleiben.

3.4. Änderungen im Einzelnen

Adriaans, Jule; Eisnecker, Phillip; Hülle, Sebastian; Klassen, J.; Liebig, Stefan; Valet, Peter (2019): „Erwartungen an Wirtschaft und Gesellschaft – Legitimation of Inequality over the Life-Span. Feldbericht und Codebuch zur zweiten Welle (LINOS-2).“ *DIW Data Documentation 97*. Deutsches Institut für Wirtschaftsforschung DIW Berlin.

https://www.diw.de/documents/publikationen/73/diw_01.c.620780.de/diw_datadoc_2019-097.pdf

Modul 1: Lebenssituation und Arbeitsplatz (bv1 - bv3)

Alle Merkmale sind subjektive Einschätzungen und bleiben erhalten.

Modul 2: Derzeitige Erwerbstätigkeit (bv4 - bv16)

bv5 – Erwerbssituation: vergrößert 3=4 5=6 10=11 (das resultierende Merkmal entspricht den Informationen im Mikrozensus-SUF 2011)

bv5_o – Erwerbssituation Sonstige: gelöscht

bv6a – Monat frühere Erwerbstätigkeit: gelöscht

bv11a_o bv11b_o bv11c bv11c_o – Berufliche Tätigkeit: Freitexte gelöscht, Berufe als ISCO-4-steller

bv13a – Anzahl der beaufsichtigten Personen: klassifiziert: 1-9 bleibt, 10-99 in 10er-Schritten, 100-199, 200 und mehr bei größeren Werten

bv14_o – Anderer Schulabschluss: gelöscht

bv15l_o – Anderer beruflicher Abschluss: gelöscht

Modul 3: Derzeitiger Arbeitsplatz (bv17-bv52)

bv40 – Tatsächliches Nettoeinkommen: Zensierung auf 10.000 €

bv41 – Tatsächliches Bruttoeinkommen: Zensierung auf 10.000 €

Modul 4: Freunde und soziale Beziehungen (bv53 – bv67)

Es werden Informationen zu drei Personen erfragt, mit denen der Befragte häufig zusammen ist. Diese Informationen sind hinreichend grob, dass sie weder das Reidentifikationsrisiko des Befragten erhöhen noch Aufschluss über die Person geben, von der die Rede ist.

Lediglich Fragen mit offenen Textangaben werden gelöscht.

bv56_1_i_o – In welcher Beziehung stehen Sie zu Person 1, Sonstige und zwar: gelöscht

bv56_2_i_o – In welcher Beziehung stehen Sie zu Person 2, Sonstige und zwar: gelöscht

bv56_3_i_o – In welcher Beziehung stehen Sie zu Person 3, Sonstige und zwar: gelöscht

Modul 5: Persönlicher Hintergrund (av61 - av71)

Gelöscht: Teil von LINOS-1 (SUF)

Modul 6: Gesellschaftliche und persönliche Werte (bv68 – bv79)

Alle Merkmale sind entweder hinreichend grob erhoben oder weisen subjektive Einschätzungen auf, so dass sie erhalten bleiben.

Modul 7: Familiäre Situation (bv80–bv112)

bv84_o– Partner: Höchster Schulabschluss; Anderer Schulabschluss, und zwar: gelöscht

bv85l_o– Partner: beruflicher Ausbildungsabschluss; Anderer Abschluss, und zwar: gelöscht

bv86 – Erwerbssituation Partner: vergrößert 3=4 5=6 10=11 (das resultierende Merkmal

entspricht den Informationen im Mikrozensus-SUF 2011)

bv86_o – Erwerbssituation Partner; Sonstige, und zwar: gelöscht

bv88a – Monat frühere Erwerbstätigkeit Partner: gelöscht

v89_o bv90a_o bv90b bv90b_o bv90c bv90c_o – Berufliche Tätigkeit: Freitexte gelöscht, Berufe als ISCO-4-steller

bv94a – Anzahl Personen im Haushalt: auf 7 zensieren

bv94b – Anzahl Personen im Haushalt, 15 Jahre und jünger: auf 5 zensieren

bv96 – Anzahl leibliche Kinder: auf 5 zensieren

bv97 – Anzahl leibliche Kinder im Haushalt: auf 5 zensieren

bv99 – Haushaltsnettoeinkommen: Zensierung auf 10.000 €

bv101j_o – Bezugsquelle Haushaltseinkommen, Sonstige Quelle, und zwar: gelöscht

bv105_o – Wahlabsicht, Andere Partei, und zwar: gelöscht

Modul 8: Angaben zur Person (bv106 – bv112)

bv107a – Geburtsmonat: gelöscht

bv107b – Geburtsjahr: Zensierung von Extremwerten auf Geburtsjahr vor 1954 und nach 1993

bv108_o – Mitglied einer anderen Religionsgemeinschaft: gelöscht

Vignetten (ev-, av-, bchoice-Merkmale)

Vignetten und Choice-Sets sind subjektive Einschätzungen und bleiben erhalten.

Metavariablen (m1 – m4)

m1 – persönliche Identifikationsnummer: wird durch eine systemfreie ID ersetzt

Merkmale zur Interviewsituation bei CAPI (bi1 – bi11)

bi1 – Interviewer-ID: wird durch eine systemfreie Interviewer-ID ersetzt

bi3a_o – Eingriff ins Interview durch Dritte bei Frage: gelöscht

bi_anmerkung – Anmerkungen zum Interview, offen Angabe: gelöscht

Generierte Merkmale

Es bleiben die zwei Gewichtungsfaktoren enthalten: Das Erstteilnahmegewicht (aweight) und das Längsschnittgewicht (bweight)

panel_consent – Einwilligung zur erneuten Kontaktierung (final)

4. Datenzugang

Die Bereitstellung der faktisch anonymisierten Daten und der Materialien zur Studie „Erwartungen an Wirtschaft und Gesellschaft“ als Scientific Use File erfolgt über das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin)

DOI SUF Erwartungen an Wirtschaft und Gesellschaft:

https://doi.org/10.25652/diw_data_S0027.1

Datensatzzitation:

Liebig, Stefan; Adriaans, Jule; Eisnecker, Philipp; Hülle, Sebastian; Klassen, Julian; Sauer, Carsten; Valet, Peter (2019): Erwartungen an Wirtschaft und Gesellschaft – Legitimation of Inequality Over the Life-Span, 2. Welle (LINOS-2). Scientific Use File. FDZ-BO am DIW Berlin. DOI:10.25652/diw_data_S0027.1

Kontakt

Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin)

Mohrenstraße 58

10117 Berlin

fdz-bo@diw.de